IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of	: HAYARDENY et al.	
Serial No.	: 10/673,733	Group Art Unit: 2161
Filed	: September 29, 2003	Examiner: Paul Kim
For	: STORAGE DISASTER RECOVERY USING A PREDICTED SUPERSET OF UNHARDENED PRIMARY DATA	

Honorable Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

APPEAL BRIEF

(1) Real Party in Interest

The subject application is owned by International Business Machines Corporation, having a place of business at New Orchard Road, Armonk, New York. The assignment was recorded in the U.S.P.T.O. on May 5, 2006, under Reel 017578, Frame 0321.

(2) Related Appeals and Interferences

None.

(3) Status of Claims

This application contains claims 1-17, 19 and 20, all of which were finally rejected in an Official Action dated August 19, 2008.

On September 25, 2008, Appellant appealed from the rejection of claims 1-17, 19 and 20 (all the claims currently pending in this application).

(4) Status of Amendments

No amendments have been made since the Official Action of August 19, 2008.

(5) Summary of Claimed Subject Matter

In the summary below and throughout this Appeal Brief, paragraph numbers refer to the published version of this application, US 2005/0081089.

One aspect of Appellant's invention, as recited in independent claim 1, provides a method for managing a data storage system that includes primary and secondary storage subsystems, including respective first and second non-volatile storage media. A system of this sort is shown in Fig. 1, wherein storage nodes A and B serve as primary and secondary storage subsystems 22 and 24 and have respective non-volatile storage media 34. A host computer 26 communicates with subsystem 22 (paragraph 0039).

The data storage system performs the following actions:

(a) A record is maintained on the secondary storage subsystem, which is predictive of locations to which data are to be written on the primary storage subsystem by a host processor. The record includes a designation of locations to which the host is expected to write in the near future. This record is shown in Fig. 2 as "bitmap 44," wherein each bit corresponds to a location. As explained in paragraph 0049, the record is referred to as the "maybe-out-of-sync" (MOOS) bitmap, and includes "locations that are not currently out of sync, but to which host 26 is expected to write in the near future."

(b) The primary storage subsystem receives data from the host processor, to be written to a specified location on the first non-volatile storage media. This function is shown at step 70 in Fig. 3 and is described in paragraph 0054.

(c) A determination is made that the specified location is not included in the record, as shown in step 74 in Fig. 3. Specifically, as explained in paragraph 0055, the primary storage subsystem determines whether the bit corresponding to the specified location is set in the copy of bitmap 44 that the primary subsystem maintains. Responsively to the determination a message is sent from the primary storage subsystem to the secondary storage subsystem so as to cause the secondary storage subsystem to update the record. Transmission of the message is shown at step 78 and described in paragraph 0056. The consequent update of the MOOS bitmap by the secondary storage subsystem is described in paragraph 0057.

(d) Sending the message causes the secondary storage subsystem to predict one or more further locations to which the host processor has not yet written the data and is expected to write the data in a subsequent write operation, and to set a number of predicted locations in the record corresponding to the one or more further locations. This predictive setting of locations (or equivalently, bits) is described in paragraphs 0055 and 0057. An example of a predictive algorithm that can be used to accomplish this step is shown in Fig. 4.

(e) Responsively to receiving the data and, when it is determined that the specified location was not included in the record, responsively to receiving an acknowledgment at the primary storage subsystem from the secondary storage subsystem indicating that the record has been updated, the system signals the host processor to indicate that the data have been stored in the data storage system. Transmission of the acknowledgment from the secondary to the primary storage subsystem is shown at step 80 in Fig. 3, and the consequent acknowledgment to the host is shown at step 82. These steps are described in paragraph 0059.

(f) The data are stored in the specified location on both the first and second non-volatile storage media. This copying is referred to as "hardening" and is described, for example, in paragraph 0054 and is also shown in step 108 of Fig. 5.

Independent claim 10 recites a method for managing a data storage system that includes primary and secondary storage subsystems, including respective first and second non-volatile storage media. A system of this sort is shown in Fig. 1, wherein storage nodes A and B serve as primary and secondary storage subsystems 22 and 24 and have respective non-volatile storage media 34. A host computer 26 communicates with subsystem 22 (paragraph 0039).

The data storage system performs the following actions:

(a) A record is maintained on the secondary storage subsystem, which is predictive of locations to which data are to be written on the primary storage subsystem by a host processor. This record is shown in Fig. 2 as "bitmap 44," wherein each bit corresponds to a location. As explained in paragraph 0049, the record is referred to as the "maybe-out-of-sync" (MOOS) bitmap, and includes "locations that are not currently out of sync, but to which host 26 is expected to write in the near future." A copy of the

record is maintained on the primary storage subsystem, which is referred to in paragraph 0049 as the "marked-on-secondary" (MOS) bitmap.

(b) The primary storage subsystem receives data from the host processor, to be written to a specified location on the first non-volatile storage media. This function is shown at step 70 in Fig. 3 and is described in paragraph 0054.

(c) A determination is made that the specified location is not included in the record, as shown in step 74 in Fig. 3. Specifically, as explained in paragraph 0055, the primary storage subsystem determines whether the bit corresponding to the specified location is set in the copy of bitmap 44 that the primary subsystem maintains. Responsively to the determination a message is sent from the primary storage subsystem to the secondary storage subsystem so as to cause the secondary storage subsystem to update the record. Transmission of the message is shown at step 78 and described in paragraph 0056. The consequent update of the MOOS bitmap by the secondary storage subsystem is described in paragraph 0057. The primary storage subsystem decides to send the message responsively to the copy of the record, i.e., depending on the corresponding bit in the MOS bitmap, as shown in step 74 and explained in paragraph 0055.

(d) Sending the message causes the secondary storage subsystem to predict one or more further locations to which the host processor has not yet written the data and is expected to write the data in a subsequent write operation, and to set a number of predicted locations in the record corresponding to the one or more further locations. This predictive setting of locations (or equivalently, bits) is described in paragraphs 0055 and 0057. An example of a predictive algorithm that can be used to accomplish this step is shown in Fig. 4.

(e) Responsively to receiving the data and, when it is determined that the specified location was not included in the record, responsively to receiving an acknowledgment at the primary storage subsystem from the secondary storage subsystem indicating that the record has been updated, the system signals the host processor to indicate that the data have been stored in the data storage system. Transmission of the acknowledgment from the secondary to the primary storage subsystem is shown at step 80 in Fig. 3, and the

consequent acknowledgment to the host is shown at step 82. These steps are described in paragraph 0059.

(f) The data are stored in the specified location on both the first and second non-volatile storage media. This copying is referred to as "hardening" and is described, for example, in paragraph 0054 and is also shown in step 108 of Fig. 5.

(6) Grounds of Rejection to be Reviewed on Appeal

Claims 1-12, 19 and 20 were rejected under 35 U.S.C. 103(a) over Yanai et al. (U.S. Patent 5,742,792) in view of Armangau et al. (U.S. Patent Application Publication 2004/0267836). Claims 13 and 14 were rejected under 35 U.S.C. 103(a) over Yanai in view of Armangau and further in view of Dunham (U.S. Patent 6,269,431). Claims 15-17 were rejected under 35 U.S.C. 103(a) over Yanai in view of Armangau and Dunham and further in view of Official Notice. Appellant believes these rejections should be reversed

(7) Argument

I. The Section 103(a) Rejection of Independent Claim 1

Appellant respectfully submits that the Examiner erred in maintaining that claim 1 is obvious over Yanai in view of Armangau.

A. The claimed invention and its benefits

Claim 1 relates to a method for managing a data storage system in which data are stored on non-volatile storage media in both primary and secondary storage subsystems. This is known as a “data mirroring” configuration, which permits the data stored on the primary subsystem to be recovered, in case of disaster, by copying data back from the secondary storage subsystem. When a host processor writes data to the primary storage subsystem, the primary storage subsystem automatically sends a copy of the data to the secondary storage subsystem. In some known mirroring configurations, for maximal data security (as explained in paragraph 0004 of the specification), the primary storage subsystem may be configured to wait for the secondary storage subsystem to confirm that it has received the data before

acknowledging to the host that the data have been stored. In this way, the host is assured that even if the primary storage subsystem crashes at the next moment, all stored data will be recoverable. This approach, known as synchronous mirroring, is safe, but slow (paragraph 0006).

For greater speed, an asynchronous approach may be used, in which the primary storage subsystem acknowledges write operations to the host immediately, and then sends copies of the data to the secondary storage subsystem in a separate, background operation. In this case, however, data written to the primary storage subsystem just before a crash are likely to be lost. Furthermore, there may be no record available to indicate which data on the primary storage subsystem were changed in the period just before the crash. Therefore, recovery may necessitate copying the entire contents of at least one storage volume in the secondary storage subsystem back to the corresponding volume in the primary storage subsystem.

The claims in the present patent application recite a novel hybrid approach, which enables low-latency write response and permits rapid recovery after failure (paragraphs 0007-0008). In this approach, the mirroring process is controlled using a metadata record that identifies storage locations that may be "out of sync" (i.e., may contain different data) on the primary and secondary storage subsystems. Storage systems known in the art have used a record of locations that are out of sync to keep track of mirroring processes. The metadata record of the present invention is different, however, in that the locations identified in the mirroring record constitute a <u>predictive superset</u> of the locations that are actually out of sync. In other words, the record contains not only the locations that are actually out of sync, but also some locations that are expected to become out of sync in the near future due to <u>host write operations that have not yet taken place</u>. The predictive record is maintained on the secondary storage subsystem, for use in subsequent disaster recovery (as explained below), but a copy of the record may be maintained on the primary storage subsystem, as well.

Upon receiving data from the host processor to be written to a specified location in the primary storage subsystem, and determining that this particular location is not included in the predictive record, the primary storage subsystem sends a message to the secondary storage subsystem. This message causes the secondary storage subsystem to

update its predictive record not only to include the specified location for this write operation, but also to predict and add to the record <u>one or more further locations to which the host processor is expected to write in a subsequent write operation but has not yet written data</u>. The primary storage subsystem then waits to signal the host processor that the data have been stored until it receives an acknowledgment from the secondary storage, indicating that the predictive record has been updated. In such cases, the data storage subsystem behaves as though it is configured for synchronous mirroring.

On the other hand, if the location specified by a given host write operation is already included in the predictive record, the primary storage subsystem may signal the host processor to indicate that the write operation has been completed without waiting for any acknowledgment from the secondary storage subsystem (paragraph 0009). In this case, the write operation can be completed very quickly, as in an asynchronous mirroring configuration. By judicious choice of the method of prediction, it is possible to predict correctly the large majority of host write locations, so that relatively few write operations will be slowed down by synchronous messaging between the primary and secondary storage subsystems. Thus, the effective write speed of the storage system will approach that of systems that use asynchronous mirroring.

When a failure occurs in the primary subsystem, the secondary subsystem can use its metadata record to determine the locations from which it should copy data back to the primary subsystem for purposes of recovery and synchronization between storage volumes. Because the metadata record contains a superset of the locations that are out of sync, it may include some locations that are actually in sync (because the host processor did not actually write to them yet). These locations will therefore be copied back unnecessarily upon recovery. The time required for recovery, however, will still be much smaller than in conventional asynchronous mirroring systems, in which entire storage volumes must typically be copied. As explained in paragraph 0010, the size of the predicted superset may be controlled so as to achieve the desired balance between write latency (which becomes shorter as the predictive superset is enlarged) and recovery time (which becomes shorter as the superset is reduced).

B. The cited art and Examiner's arguments

Yanai describes a remote data mirroring system, which stores data received from a host computer on a primary data storage system and copies the data to a secondary data storage system for providing a back-up copy (col. 2, lines 30-38). Both the primary and secondary data storage systems maintain a table of the validity of data in each storage system (col. 11, lines 10-25). (An index is also maintained of pending format changes - col. 11, lines 25-30.) When a host computer writes data to the primary storage system, it sets both primary and secondary write pending bits in the table (col. 11, lines 31-34). These bits indicate that the host computer has written data to the cache, and that the controllers of the primary and secondary storage systems must now write the data to the respective storage devices, whereupon the bits will be reset (col. 11, lines 37-43).

The Examiner cited the above passages, in col. 11, as purportedly teaching "a record... which is predictive of locations to which data are to be written to the primary storage subsystem by a host processor," which includes "a designation of locations to which the host is expected to write in the near future." The cited passages, however, relate only to locations to which the host computer has already written data. There is nothing in Yanai's record that is even remotely predictive of locations to which a host is expected to write, as recited in claim 1. Armangau fails to remedy this deficiency.

Armangau describes a snapshot copy facility that determines the difference between an older snapshot copy and a younger snapshot copy (abstract). This facility is used in copying modifications from a primary file system to a secondary file system using "delta chunks" (paragraphs 0079-0081). Armangau mentions hosts in passing (paragraph 0068), but provides no teaching or suggestion whatsoever that these hosts might write anything to the file servers that hold his file systems. In other words, not only does Armangau fail to teach or suggest a predictive record of locations to which a host processor is expected to write: He does not relate to any sort of record at all of locations of host writes. Therefore, a person of ordinary skill in the art could not possibly have been motivated to incorporate teachings from Armangau into the records of host writes that Yanai maintains.

Furthermore, even if Armangau were considered, for the sake of argument, to be relevant to the invention of claim 1, he still fails to teach or suggest <u>predicting further locations to which data are to be written, or setting a number of predicted locations in any sort of record</u>. In regard to this element of claim 1, the Examiner cited paragraphs 0099 and 0243 in Armangau. Paragraph 0099 describes the use of two bit-map tables, indicating blocks in a primary file system volume that have been modified (i.e., new data <u>have already been written</u> to these blocks), and which must therefore be copied on a priority basis to a save volume. Paragraph 0243 refers to a "meta bit map," which indicates whether or not each allocated block of storage is valid or not. Armangau does not explicitly define what he means by "valid," but he uses the term "invalid" to refer to blocks whose contents need not be saved (paragraph 0241). The indication of "validity" provided by the meta bit map therefore has nothing to do with either writing or prediction.

In the "Response to Arguments" (page 9 in the Official Action), the Examiner referred back to the same passages as he cited previously, and stated that "Accordingly, it is noted that Yanai discloses an invention wherein the host computer (i.e. the host) is able to set bits indicating a write or copy function to either a primary or secondary data storage (i.e. the location)." Yes, and as noted above, this function takes place "when a host computer writes data" to the primary storage system (col. 11, lines 31-34). This is no more than a conventional marking of locations that are out of sync, as explained above. <u>The Examiner has not pointed to any teaching or suggestion, in either Yanai or Armangau, of a predictive record, holding locations to which the host processor has not yet written data and is expected to write data in a subsequent write operation</u>, as recited in claim 1.

C. Conclusion

MPEP 2143.03 makes clear that: "To establish *prima facie* obviousness of a claimed invention, all the claim limitations must be taught or suggested by the prior art. *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974)." Although the recent Supreme Court decision in *KSR International Co. v. Teleflex Inc. et al*, 550 U.S. __ (2007) has relaxed the "TSM" test for combining references, it made no change in the

requirement that all claim limitations must be taught or suggested by the prior art. The Examiner has failed to meet this burden.

Because Yanai and Armangau together fail to teach or suggest all the limitations of claim 1, this claim is patentable over the cited art. Therefore, the rejection of claim 1 should be reversed.

II. The Section 103(a) Rejection of Independent Claim 10

Appellant respectfully submits that the Examiner erred in maintaining that claim 10 is obvious over Yanai in view of Armangau.

Independent claim 10 is similar to claim 1, with the added feature that the primary storage subsystem maintains a copy of the record on the secondary storage subsystem and uses the copy in determining whether to send the message. Claim 10 is therefore patentable, as well, for the reasons stated above with respect to claim 1: The cited art fails to teach or suggest any sort of predictive record, let alone the novel method of maintaining and using the record that is recited in claim 10.

Furthermore, in rejecting claim 10 together with claim 1, the Examiner failed to mention or relate at all to the added feature in claim 10 of maintaining a copy of the predictive record on the primary storage subsystem, or deciding at the primary storage subsystem to send a message to the secondary storage subsystem responsively to the copy of the record. This feature is neither taught nor suggested by the cited art. Therefore, claim 10 is independently patentable notwithstanding the patentability of claim 1.

Because Yanai and Armangau together fail to teach or suggest all the limitations of claim 10, this claim is patentable over the cited art. Therefore, the rejection of claim 10 should be reversed.

III. The Section 103(a) Rejection of Claim 3

Appellant respectfully submits that even if independent claim 1 were conceded to be unpatentable over Yanai in view of Armangau, the cited references still do not teach or suggest the added elements of dependent claim 3.

Claim 3 depends from claim 2, which depends from claim 1, and relates (based on claim 2) to embodiments in which data are copied synchronously from the primary to the secondary storage subsystem. Claim 3 recites that, in contrast to claim 1, when it is determined that the specified location is included in the predictive record, the data are copied from the primary storage subsystem asynchronously, without updating the record with respect to the specified location. Thus, depending upon whether the location is or is not included in the predictive record, the primary storage subsystem decides whether to copy the data synchronously or asynchronously to the secondary storage subsystem, and whether or not to update the record.

The cited art neither teaches nor suggests a method in which data are sometimes copied synchronously, and sometimes asynchronously, on the basis of a given record. In rejecting claims 2 and 3, the Examiner cited Yanai's references to a synchronous mode (col. 15, lines 35-39) and to the ability of the host computer to request either synchronous or asynchronous data transfer between primary and secondary data storage system controllers (col. 2, lines 58-65). Appellant will readily acknowledge that data storage systems known in the art prior to the present invention could be configured for either synchronous or asynchronous operation under host control. There is no teaching or suggestion in the cited art, however, that a <u>primary storage subsystem might itself decide, for each write location specified by the host, whether to copy the data synchronously or asynchronously</u>, depending on a record that the primary storage subsystem itself maintains and updates, as recited in claim 3.

Therefore, claim 3 is independently patentable over the cited art, notwithstanding the patentability of independent claim 1.

IV. The Section 103(a) Rejection of Claim 5

Appellant respectfully submits that even if claims 1-3 were conceded to be unpatentable over the cited art, the cited references still do not teach or suggest the added elements of dependent claim 5. This claim depends from claim 3 and recites that the primary storage subsystem maintains a copy of the predictive record. Upon determining that the location specified by a host write operation <u>is included in the</u>

record, the primary storage subsystem indicates to the host processor that the data have been stored without waiting to receive the acknowledgment from the secondary storage subsystem. This claim is the "flip side" to claim 1, which states that when the location is not included in the record, the primary storage subsystem waits for an acknowledgment from the secondary storage subsystem before signaling the host processor that the data have been stored.

In rejecting this claim, the Examiner referred to Yanai's "semi-synchronous mode," in which the "data storage system containing the primary (R1) informs the host that an I/O sequence has successfully completed without waiting for the data storage system containing the secondary (R2) volume to acknowledge that it has received and checked the data" (col. 3, lines 43-48). This is no more than a description of a modified version of asynchronous data mirroring. In Yanai's semi-synchronous mode, all host write operations are treated in this fashion. Yanai neither teaches nor suggests that a primary storage subsystem might consult some record of storage locations, and decide on the basis of this record when to acknowledge the host write without waiting for acknowledgment from the secondary storage subsystem, as recited in claim 5.

Therefore, claim 5 is independently patentable over the cited art, notwithstanding the patentability of claims 1-3.

V. The Section 103(a) Rejection of Claim 12

Appellant respectfully submits that even if independent claim 10 were conceded to be unpatentable over Yanai in view of Armangau, the cited references still do not teach or suggest the added elements of dependent claim 12.

Claim 12 depends from claim 11, which depends from claim 10, and recites what happens when the primary storage subsystem sends a message to the secondary storage subsystem: Both the predictive record on the secondary storage subsystem and the copy of the record on the primary storage subsystem are modified by adding a plurality of locations, including the specified location, to both the record and the copy of the record. In other words, one message from the primary to the secondary storage

subsystem, indicating that the current write location is not in the record, causes multiple locations to be added to the record by both subsystems.

The cited art does not teach or suggest anything that even remotely approximates this method. In rejecting claim 12, the Examiner referred to Yanai's description of setting the primary and secondary write pending bits 104 and 106 when data are written to cache (col. 11, lines 31-34). It is clear from the context, however, that only a single bit (M1) is set in reference to the primary storage system and another single bit (M2) is set in reference to the secondary storage system (col. 11, lines 35-38). Each bit can represent no more than a single storage location – corresponding to the "specified location" of claim 12, which is the location specified by the host processor. Yanai does not even hint that a plurality of bits, or a plurality of locations, should be set in both the primary and secondary records, as required by claim 12.

Therefore, claim 12 is independently patentable over the cited art, notwithstanding the patentability of claim 10.

VI. The Section 103(a) Rejection of Claim 16

Appellant respectfully submits that even if independent claim 10 were conceded to be unpatentable over Yanai in view of Armangau, the cited references still do not teach or suggest the added elements of dependent claim 16. This claim depends from claim 13, which depends from claim 10 and relates to selecting locations to be removed from the predictive record. Claim 13 was rejected over Yanai and Armangau in view of Dunham (U.S. Patent 6,269,431). In rejecting claim 16, the Examiner also relied on Official Notice in addition to the cited references, and then went on to cite Menon et al. (U.S. Patent 6,397,229) as documentary evidence in support of the Official Notice (page 10 in the Official Action).

Claim 16 recites grouping entries in the predictive record in generations on both the primary and secondary subsystems according to the order in which the entries were added to the record, and then determining at the primary subsystem that all the entries in one of the generations may be removed from the record. This approach to managing the predictive record is useful in reducing the communication burden between the primary

and secondary storage subsystems, since it permits multiple different entries to be removed using a single message.

In regard to claim 16, the Examiner took the position that "the grouping of entries added to the record have been construed as versions" (page 10 in the Official Action) and that it would have been obvious to discard such a "generation or version" using a "batch method" (page 8). The Examiner has given no rationale as to why entries of locations in a predictive record might be considered "versions," nor why the same "versions" should be grouped and then removed from both primary and secondary subsystems, as recited in claim 16. Although Menon describes setting and clearing bits in a storage bit map, (col. 4, line 40 – col. 5, line 6), he makes no mention of "versions" of any sort and does not even hint that the bits in his bit map might be cleared more than one at a time.

The conclusion of the Examiner's argument in this regard is that Menon's "bits could be grouped and removed from the records when the sectors are backed up to the tape drive." This is no more than a conclusory statement, <u>without support in Menon or any other rational underpinning</u>, except for hindsight from the present patent application. As such, the Examiner has failed to meet the criteria set forth by the Supreme Court in *KSR* for supporting a finding of *prima facie* obviousness.

Therefore, claim 16 is independently patentable over the cited art, notwithstanding the patentability of claims 10 and 13.

VII. The Section 103(a) Rejection of Claim 17

Appellant respectfully submits that even if claim 10 were conceded to be unpatentable over Yanai in view of Armangau, and claim 13 were conceded to be unpatentable over these references together with Dunham, the cited references still do not teach or suggest the added elements of dependent claim 17.

Claim 17 recites that the very same message from the primary storage subsystem that instructs the secondary storage subsystem to update its record by setting a number of predicted locations in the record (as recited in claim 10) is also used in instructing the secondary storage subsystem to remove one or more locations from the record. As in

the case of claim 16, this technique helps to reduce the communication burden between the primary and secondary storage subsystems.

The Examiner rejected claim 17 over Yanai, Armangau and Dunham and further in view of Official Notice (page 8 in the Official Action), and then went on to cite Menon as documentary evidence in support of the Official Notice (page 11 in the Official Action). In his Official Notice with respect to claim 17, the Examiner did no more than to restate the limitations of the claim itself, without giving any rationale at all as to why these limitations might be obvious, particularly in the context of the claimed invention. In regard to Menon, the Examiner cited nothing more relevant that the basic statement that when "a backup/write request for a sector is sent and processed, the bit associated with the sector is changed."

Claim 17, however, says something much more specific: that <u>the same message causes the secondary storage subsystem both to predictively add locations to the record (per claim 10) and to remove locations from the record</u>. The Examiner has not given any sort of rationale as to why this novel sort of messaging should be considered in any way obvious over the prior art.

Therefore, claim 17 is independently patentable over the cited art, notwithstanding the patentability of claims 10 and 13.

VIII. The Section 103(a) Rejection of Claim 20

Appellant respectfully submits that even if independent claim 1 were conceded to be unpatentable over Yanai in view of Armangau, the cited references still do not teach or suggest the added elements of dependent claim 20.

Claim 20 depends from claim 1 and recites that an object-based storage technique is used in recording the locations to which data are written in the data storage system of the claim. As explained in the specification (paragraph 0067), in object-based storage, the control unit of the storage system is aware of the association between storage locations and logical objects, such as HTML pages. The message sent from the primary storage subsystem, in response to a host write, causes the secondary storage

subsystem to set a number of predicted locations in the record based on a logical connection between storage objects.

In rejecting claim 20, the Examiner relied on a passing reference to "logical volumes" in Yanai (col. 3, lines 25-29). Logical volumes are well known in the art as a way of allocating and managing space in a data storage system. Yanai makes no mention or suggestion of object-based storage, as this term is defined in the present patent application and is understood in the art of storage systems. Furthermore, even if Yanai's logical volumes were considered to be "storage objects," he says nothing about "logical connections" between such objects. Therefore, he cannot possibly be taken to suggest selecting further locations to be added to a predictive record based on such logical connections, as required by claim 20.

Thus, claim 20 is independently patentable over the cited art, notwithstanding the patentability of independent claim 1.

Summary

For the foregoing reasons, it is submitted that the Examiner's rejection of claims 1-17, 19 and 20 was erroneous. Reversal of the Examiner's decision is respectfully requested.

Respectfully submitted,

By: /Suzanne Erez/
Suzanne Erez
Reg. No. 46,688
Phone No. (972) 4-829-6069

Date: 10 November 2008
IBM Corporation
Intellectual Property Law Dept.
P. O. Box 218
Yorktown Heights, New York 10598

APPENDIX A

Claims under appeal:

1. A method for managing a data storage system that includes primary and secondary storage subsystems, including respective first and second non-volatile storage media, the method comprising:

maintaining a record on the secondary storage subsystem, which is predictive of locations to which data are to be written on the primary storage subsystem by a host processor, the record including a designation of locations to which the host is expected to write in the near future;

receiving at the primary storage subsystem, from the host processor, the data to be written to a specified location on the first non-volatile storage media;

making a determination that the specified location is not included in the record, and responsively to the determination sending a message from the primary storage subsystem to the secondary storage subsystem so as to cause the secondary storage subsystem to update the record,

wherein sending the message causes the secondary storage subsystem to predict one or more further locations to which the host processor has not yet written the data and is expected to write the data in a subsequent write operation, and to set a number of predicted locations in the record corresponding to the one or more further locations;

signaling the host processor that the data have been stored in the data storage system responsively to receiving the data and, upon having made the determination that the specified location was not included in the record, responsively to receiving an

acknowledgment at the primary storage subsystem from the secondary storage subsystem indicating that the record has been updated; and

storing the data in the specified location on both the first and second non-volatile storage media.

2. The method according to claim 1, wherein sending the message comprises copying the data synchronously from the primary storage subsystem to the secondary storage subsystem.

3. The method according to claim 2, wherein storing the data comprises, upon making the determination that the specified location is included in the record, copying the data from the primary storage subsystem to the secondary storage subsystem asynchronously, without updating the record with respect to the specified location.

4. The method according to claim 3, wherein copying the data comprises transmitting the data between mutually-remote sites over a communication link between the sites.

5. The method according to claim 3, wherein maintaining the record comprises maintaining a copy of the record on the primary storage subsystem, and wherein signaling the host processor comprises, upon making the determination that the specified location is included in the record, indicating to the host processor that the data have been stored without waiting to receive the acknowledgment from the secondary storage subsystem.

6. The method according to claim 1, wherein copying the data comprises creating a mirror on the secondary storage subsystem of the data received by the primary storage subsystem.

7. The method according to claim 6, and comprising, upon occurrence of a failure in the primary storage subsystem, configuring the secondary storage subsystem to serve as the primary storage subsystem so as to receive further data from the host processor to be stored by the data storage system.

8. The method according to claim 6, and comprising, upon recovery of the system from a failure of the primary storage subsystem, conveying, responsively to the record, a portion of the data from the secondary storage subsystem to the primary storage subsystem for storage on the primary storage subsystem.

9. The method according to claim 1, wherein maintaining and updating the record comprise marking respective bits in a bitmap corresponding to the locations to which the data are to be written on the first and second non-volatile storage media.

10. A method for managing a data storage system that includes primary and secondary storage subsystems, including respective first and second non-volatile storage media, the method comprising:

maintaining a record on the secondary storage subsystem, which is predictive of locations to which data are to be written on the primary storage subsystem by a host processor, wherein maintaining the record comprises maintaining a copy of the record on the primary storage subsystem;

receiving at the primary storage subsystem, from the host processor, the data to be written to a specified location on the first non-volatile storage media;

making a determination that the specified location is not included in the record, and responsively to the determination sending a message from the primary storage subsystem to the secondary storage subsystem so as to cause the secondary storage subsystem to update the record, wherein sending the message comprises deciding at the primary storage subsystem to send the message responsively to the copy of the record, and

wherein sending the message causes the secondary storage subsystem to predict one or more further locations to which the host processor has not yet written the data and is expected to write the data in a subsequent write operation, and to set a number of predicted locations in the record corresponding to the one or more further locations;

signaling the host processor that the data have been stored in the data storage system responsively to receiving the data and, upon having made the determination that the specified location was not included in the record, responsively to receiving an acknowledgment at the primary storage subsystem from the secondary storage subsystem indicating that the record has been updated; and

storing the data in the specified location on both the first and second non-volatile storage media.

11. The method according to claim 10, wherein sending the message comprises modifying both the record and the copy of the record responsively to the specified location.

12. The method according to claim 11, wherein modifying both the record and the copy of the record comprises adding a plurality of locations, including the specified location, to both the record and the copy of the record.

13. The method according to claim 10, wherein maintaining the copy of the record comprises selecting one or more locations, other than the specified location, to be removed from the record, and instructing the secondary storage subsystem to remove the one or more locations from the record, so as to limit a size of the record.

14. The method according to claim 13, wherein storing the data comprises copying the data to be stored in the one or more locations from the primary storage subsystem to the secondary storage subsystem, and wherein selecting the one or more locations comprises receiving a return message from the secondary storage subsystem indicating that the secondary storage subsystem has received the copied data, and selecting the one or more locations to be removed from the record responsively to the return message.

15. The method according to claim 13, wherein selecting the one or more locations comprises identifying the locations at which the first and second non-volatile storage media contain substantially identical data, and selecting for removal one of the identified locations that was least-recently added to the record.

16. The method according to claim 13, wherein sending the message comprises adding one or more entries to both the record and the copy of the record responsively to the specified location, and grouping the entries added to the copy of the record and the record in generations according to an order of adding the entries to the

records, and wherein selecting the one or more locations comprises determining at the primary subsystem that all the entries in one of the generations may be removed from the record.

17. The method according to claim 13, wherein instructing the secondary storage subsystem to remove the one or more locations comprises appending an instruction to the message sent from the primary storage subsystem to the secondary storage subsystem.

19. The method according to claim 1, wherein the one or more further locations comprise a predetermined number of consecutive locations in proximity to the specified location.

20. The method according to claim 1, wherein maintaining the record comprises recording the locations to which the data are written using an object-based storage technique, and wherein the one or more further locations are chosen based on a logical connection between storage objects.

APPENDIX B – EVIDENCE

None presented.

APPENDIX C – RELATED PROCEEDINGS

None.